Research Article

The Fault Diagnosis of Rolling Bearing Based on Improved Deep Forest

Xiwen Qin ^{1,2}, Dingxin Xu ¹, Xiaogang Dong ¹, Xueting Cui ¹ and Siqi Zhang ¹

¹School of Mathematics and Statistics, Changchun University of Technology, Changchun 130012, China

²Graduate School, Changchun University of Technology, Changchun 130012, China

Correspondence should be addressed to Xueting Cui; 201611010@stu.ccut.edu.cn

Received 30 March 2021; Accepted 7 June 2021; Published 24 June 2021

Academic Editor: Arturo Garcia-Perez

Copyright © 2021 Xiwen Qin et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Rolling bearing fault diagnosis is a meaningful and challenging task. Most methods first extract statistical features and then carry out fault diagnosis. At present, the technology of intelligent identification of bearing mostly relies on deep neural network, which has high requirements for computer equipment and great effort in hyperparameter tuning. To address these issues, a rolling bearing fault diagnosis method based on the improved deep forest algorithm is proposed. Firstly, the fault feature information of rolling bearing is extracted through multigrained scanning, and then the fault diagnosis is carried out by cascade forest. Considering the fitting quality and diversity of the classifier, the classifier and the cascade strategy are updated. In order to verify the effectiveness of the proposed method, a comparison is made with the traditional machine learning method. The results suggest that the proposed method can identify different types of faults more accurately and robustly. At the same time, it has very few hyperparameters and very low requirements on computer hardware.

1. Introduction

Rolling bearing is an important basic device in mechanical equipment which has been widely used in wind power generation group, high-speed electric multiple unit (EMU), computerized numerical control (CNC) machine tools, and other equipment [1]. Rolling bearing is the core component of rotating machinery, whose failure will result in huge economic losses and threaten personal and property safety [2]. Therefore, it is necessary to accurately grasp the running status of rolling bearing, timely maintain the damaged parts, and prevent them from evolving into a greater threat. Accurately and effectively identifying the types of bearing fault and ensuring the normal operation of mechanical equipment are essential to improve the reliability of the system.

Vibration signals of rolling bearing fault are usually nonstationary and nonlinear [3, 4]. The early bearing fault identification technologies are mainly based on time domain, frequency domain, and time-frequency signal analysis methods [5–7]. In general, the fault features such as kurtosis, coefficient of variation, energy entropy, information

entropy, and power spectrum entropy are extracted from the original signal, and then the fault identification is carried out by combining the classification algorithm. In terms of traditional fault diagnosis methods, the more commonly used time-frequency analysis methods include wavelet transform [8], empirical mode decomposition (EMD) [9], and variational mode decomposition (VMD) [10]. Zhao proposed a rolling bearing fault diagnosis method that combines wavelet packet decomposition (WPD) and multiscale permutation entropy (MPE). The vibration signals of rolling bearing in different states are decomposed into a group of subfrequency signals by using WPD, and then the average MPE value of each subfrequency signal is calculated as the input feature vector, and the fault modes of rolling bearings were identified by the hidden Markov model (HMM) [11]. Zhang proposed an automatic fault diagnosis method for rolling bearings based on lifting wavelet packet transform (LWPT), sample entropy, and classifier integration. And the construction of wavelet function is not based on Fourier transform but is obtained in the time domain. At the same time, considering the unstable accuracy of single classifier,

an ensemble system which integrates back propagation neural network (BPNN), radial basis function neural network (RBFNN), and Elman neural network (ElmanNN) is proposed to reduce the impact of initial parameters on the performance of the classifier [12]. Compared with the traditional wavelet transform, the lifting wavelet packet transform has the advantages of the flexibility of constructing the wavelet function, less computation, and less memory. Ensemble empirical mode decomposition (EEMD) [13] is an improved version of EMD. As a typical representative of the adaptive method for dealing with nonlinear and nonstationary data, EEMD has been widely applied in the field of fault diagnosis [14, 15]. VMD is developed on the basis of EMD, which has a solid mathematical theoretical basis and has been proved to be superior to other adaptive data decomposition methods. It is widely used in fault diagnosis [16]. Chen et al. proposed the fault diagnosis of rolling bearing based on VMD and support vector machine (SVM), which significantly improved fault identification accuracy through multiscale fractal dimension and multiscale energy calculation features [17].

In addition, some shallow learning methods are closely combined with intelligent optimization algorithms for fault diagnosis. Dai proposed a fault identification method based on KICA-RBF [18]. In this model, it is important to use kernel independent component analysis (KICA) to fuse multiple signals to eliminate noise, and the genetic algorithm is used to optimize the parameters of radial basis function (RBF), thus the accuracy is improved. Zhao and Deng et al. improved a variety of optimization algorithms and proposed a data-driven feature extraction method—fitting curve derivative method of maximum power spectrum density (FDMPD)—and combined with the kernel extreme learning machine (KELM) and weight application to failure times (WAFT), which can effectively realize the prediction of the remaining service life of rolling bearings [19–22]. LV et al. proposed an improved particle swarm optimization (PSO) algorithm to optimize parameters of support vector machine for fault diagnosis of rolling bearings [23]. The PSO is improved by introducing dynamic inertia weight, global neighborhood search, population shrinkage factor, and particle mutation probability. The method solves the problem of blind selection of kernel function parameter and penalty factor parameter of SVM. Experimental results showed that the classification effect is more stable. Shallow learning algorithms require manual participation in the construction of feature engineering and have poor ability of learning representation, while deep learning can effectively model high-level abstraction of data due to its powerful nonlinear representation capability [24]. In recent years, with the development of artificial intelligence, deep learning has made breakthrough progress. The cross-domain application of deep learning in fault diagnosis has aroused great interest and achieved remarkable results [25–28]. Zhong et al. used EEMD to decompose intermittent fault signals into multiple intrinsic modal functions (IMFs), combined with Pearson correlation coefficient for feature optimization, and deep belief network (DBN) is used for fault diagnosis [29]. Guo et al. proposed a bearing fault diagnosis method

based on hierarchical learning rate adaptive deep convolutional neural network and achieved satisfactory results [30]. Xu et al. proposed a bearing fault diagnosis method based on convolutional neural network (CNN) and random forest, which took two-dimensional images of continuous wavelet transform as input [31]. The multilevel features containing local and global information are used to diagnose bearing faults. The research indicated that this method is superior to the base deep learning method. Although the fault diagnosis method based on deep neural network (DNN) is powerful, due to the complexity of its model, a large amount of training data and its learning performance are excessively dependent on parameter optimization, which limits its applicability. In 2017, Zhou et al. proposed a method different from deep learning called multigrained cascade forest (gcForest), which generates a deep forest with cascade structure for representation learning, and it is regarded as a decision tree ensemble approach [32]. gcForest is easier to analyze theoretically than DNN. Some scholars have explored its application in the field of fault diagnosis. Hu et al. proposed a collaborative method combining deep Boltzmann machine with multigrained scanning forest integration, which effectively solved the problem of industrial fault diagnosis under big data [33]. Liu et al. applied deep forest for the first time in the end-to-end intelligent diagnosis of hydraulic turbine faults [34]. Considering the diversity of the cascade forest classifiers and the classification performance of each classifier, this paper proposes an improved deep forest algorithm. The mechanism changes the cascading mode based on the output results of the multigrained scanning stage and replaces the classifier through the cascade stage to increase the diversity and improve the performance of the classifier.

The main contributions of this paper are summarized as follows:

- (1) The classification is based on the original vibration signal data, which is different from most existing literatures in which they first extract the features and then classify them. The interference of human factors to feature window and feature type is avoided.
- (2) An improved deep forest algorithm is proposed, the idea of heterogeneous integration is introduced, and the cascade mode is changed to reduce the loss of sample information, which further improved the accuracy of fault diagnosis of gcForest.

The rest of this paper is organized as follows. Section 2 is the basic principle of deep forest and its improved algorithm; in Section 3, we give the results of the empirical analysis. Section 4 is the conclusion.

2. gcForest

GcForest, also known as deep forest (DF), is a supervised ensemble learning algorithm based on decision tree, which mainly consists of two parts: multigrained scanning (MGS) and cascade forest [32]. MGS solves the problem of high-dimensional input and enhances the difference of input feature. Cascade forest can improve the classification ability

of input features by simulating the structure of DNN for representation learning.

2.1. Multigrained Scanning. The multigrained scanning is mainly inspired by the CNN. The key point of the CNN is that different enhancement features can be obtained through the convolutional kernel of different sizes. The MGS just draws on this idea to enhance the cascade forest [35]. Multigrained scanning is mainly used to locally sample the original data through the sliding window so as to obtain multiple feature instances of different dimensions. The process is described as follows: the input sample size be S dimension. The sliding window size is K dimension. The sliding step size is L , and M represents the number of generated feature vectors, and then we have the following equation:

$$M = \frac{(S - K)}{L + 1}. \quad (1)$$

After the multigrained scanning, each sample subset is input to a random forest and a completely random forest (CRF) for training so that each forest can obtain the feature vector of $M * C$, where C is the number of categories. Finally, $M * C * 2$ -dimensional features can be obtained as the output of the multigrained scanning structure. As shown in Figure 1, it is a multigrained scanning process. Assuming that the raw data have 400-dimensional feature, the size of the sliding window is 100-dimensional; after sliding scanning, it will produce 301 feature vectors. If this is a three-class problem, each forest will produce 301 three-dimensional class vectors. Finally, 1806-dimensional transformed feature vector is obtained. Similarly, for 400-dimensional input data, sliding windows of 200-dimensional and 300-dimensional will generate 1206-dimensional and 606-dimensional feature transformed vectors, respectively.

2.2. Cascade Forest. The cascade forest stage embodies the process of deep learning through the hierarchical representation learning of features. Each level in the cascade forest corresponds to a different scanning granularity. The latter level receives the feature information from the previous level. The feature information is transmitted to the next level after processing at this level. Each level takes as its input an eigenvector that connects the original input to the output from the previous level [32, 35]. As shown in Figure 2, each layer of the cascade contains two random forests and two completely random forests, each forest is composed of multiple decision trees, and each tree randomly selects $\sqrt{d}$ features from d input features as candidate features. The dividing standard of node splitting in decision tree is to select the feature with the best Gini value as the root node. Completely random forest in the choice of decision tree node split is random; each leaf node until only the similar samples has stopped growing.

The classification results are obtained by the class vectors distribution of the leaf nodes of the decision tree. And then you take the average of all the trees to get an estimate of the

class distribution. In order to reduce the risk of overfitting, the cross-validation method is used during training.

2.3. Overall Process of Deep Forest. Combining the multigrained scanning and the cascade forest, the overall flow of the deep forest is obtained, as shown in Figure 3. Suppose that there is original input of 400 features, three sliding windows with length of 100-, 200-, 300-dimensional are used for multigrained scanning, respectively [36]. The sliding step length is 1. As stated in Section 2.1, the feature vectors after MGS are input to a RF and a CRF for training, and then the eigenvectors of 1806-dimensional, 1206-dimensional, and 606-dimensional are obtained, respectively, and used for training the first level of the cascade forest. Taking the 100-dimensional sliding window as an example, suppose there are three classes. After the 1806-dimensional features are trained by four forest classifiers. Then, it is connected to the 1806-dimensional feature vector obtained after scanning transformation, so the first level obtains 1818-dimensional feature vectors, which is the input to the second level. Similarly, the second level obtains 1218-dimensional feature vector after concatenating, which is used as the input of the third-level training. The third level generates 618-dimensional vectors, which is used as the input of the next level. The above process is repeated until there is no significant performance gain, and the training process is stopped.

2.4. Improved Deep Forest Model. DF is an ensemble learning algorithm. If you want to build an ensemble with strong generalization ability, it should be “good but different” for individual learners. Layer-by-layer training of cascade forest can enhance the representation ability of feature information, and it is very important for DF ensemble learning to adopt different classifiers for each layer.

In this study, cascade forest based on multiple heterogeneous classifiers is proposed, and the classifier of each hierarchy is set to RF, ET, XGBoost, and LightGBM. The change is illustrated in Figure 4. The combination of multiple different types of forest classifiers could fully learn the feature information of the input feature vectors and improve the overall performance of the model.

The most important characteristics of RF are sample randomness and feature randomness. By extracting different training sets and randomly extracting features for training, the difference between classification models will be increased, which can effectively avoid the overfitting problem. At the same time, the parallel computing mechanism of the RF algorithm greatly reduces the training time of data [37].

ET is the abbreviated form of the extremely randomized trees model. Each time, all the samples are used for training, and features are randomly selected. Since splitting is random, the results obtained by ET are better than those obtained by RF to some extent [38].

XGBoost (eXtreme gradient boosting) is a supervised learning algorithm based on decision tree, which is proposed on the basis of gradient boosting decision tree (GBDT) [39]. The XGBoost generates a new tree by iteration. Compared with GBDT, XGBoost adds regularization term to the loss

FIGURE 1: The multigrained scanning structure.

FIGURE 2: Cascade forest structure.

FIGURE 3: Overall flow of deep forest.

function to control the complexity of the model. XGBoost supports linear classification problems and draws on the practice of RF by supporting column sampling, which not only reduces overfitting but also reduces computation, which is a feature of XGBoost that is different from GBDT.

XGBoost has been widely used because of its advantages of parallel processing, simple model structure, small computation amount, and high accuracy [40].

LightGBM is an improvement of GBDT, mainly to solve the problem of the decline of training efficiency when the

FIGURE 4: Improved deep forest.

GBDT algorithm is dealing with large amount of data. LightGBM improves the GBDT algorithm from two technical aspects as follows. (1) To solve the problem of large amount of data, the improved method is gradient-based one-side sample (GOSS). The gradient is calculated by sampling the samples, the large gradient data are retained, and the small gradient data are randomly sampled to reduce the amount of data used by the samples; (2) The Leaf-wise splitting method with depth limitation is used to replace the traditional Leaf-wise splitting method. Each time, the Leaf splitting with the largest splitting gain is found to generate a more complex decision tree, which can reduce the error and improve the accuracy of the algorithm. These two technical methods of LightGBM greatly reduce the time cost and accelerate the training process, and a large number of experimental studies have shown that LightGBM has even better performance in terms of accuracy. Therefore, XGBoost and LightGBM are selected in this paper to replace the two original classifiers in the cascade forest [41].

On the other hand, in the gcForest, the probability feature and the original input feature can be serially integrated into the input vector to effectively prevent overfitting. However, as the model depth increases, this sparse connection structure may result in a large amount of information being discarded, which may hinder the diversity of the integration. This study is inspired by the model of dense cascade forest [42]; we improve the gcForest. As shown in Figure 4, for each level of the cascade, a sublevel is added. In the first level of the original cascade, there are three sublevels called $Level1_A$, $Level1_B$, and $Level1_C$. The additional sublevels are created by connecting all the features together. Taking Section 2.3 as an example, this feature will be 3618-dimensional; then the characteristics are input into the classifier to obtain the probability class vector, which is concatenated with 3618-dimensional features as $Level1_A$ level. The original 1806-dimensional, 1206-dimensional, and 606-dimensional features are concatenated in $Level1_B$, $Level1_C$, and $Level1_D$, respectively. The process at the second level is similar to that at the first level, in which cascading all the features retains more information of the original sample. In the experiment, we find that this structure makes the training process more stable.

3. Empirical Analysis

3.1. Data Sources. The data in this study come from the public bearing data set of Case Western Reserve University [43] in the United States. The data were collected from a motor, a torque sensor, and a power tester. The experimental platform is shown in Figure 5. The fault of the bearing is manufactured by the electric spark technology. The motor load is from 0 to 3 horsepower, and the corresponding motor speed is 1797 to 1730 rpm. The vibration signal includes the normal data, the drive end acceleration data, the fan end acceleration data, and the base data. This paper only collects the fault data of the drive end for analysis, and the sampling frequency is set to 12 kHz. Four different fault diameters were introduced for inner raceway, outer raceway, and rolling element. The fault diameter ranges from 0.007 inches to 0.028 inches. Due to the lack of data with fault diameter of 0.028 inches in some types of data, only 0.007 inches, 0.014 inches, and 0.021 inches of data are retained. Rolling bearing can be divided into four kinds of condition: normal state (NOR), the inner race fault (IRF), ball fault (BF), and outer race fault (ORF). Figure 6 shows three types of bearing fault. Partial vibration signals of the four conditions are shown in Figure 7.

In this study, we design two experiments for fault diagnosis of rolling bearing. In Experiment 1, the collected data are divided into four categories, including normal state, inner race fault, rolling fault, and outer race fault. The data under each condition are 1460,000 data points. In Experiment 2, the fault data are further divided into three different fault diameters. There are three fault states in each fault diameter, plus the normal state. Therefore, the data are divided into 10 categories, and the sampling point of each category is 480000. Each type of data set is randomly divided into training set and test set in an 8:2 ratio.

3.2. Performance Evaluation and Parameters Setting. In order to evaluate the generalization ability of the model, the 8-fold cross-validation method is adopted. Accuracy and macroaverage are used to evaluate the performance of the algorithm. Accuracy is the proportion of correctly predicted

FIGURE 5: Rolling bearing fault test bench.

FIGURE 6: Three kinds of fault state: (a) inner race fault; (b) outer race fault; (c) ball fault.

samples to the total predicted samples, which represents the average degree of the classification performance of the whole model. The macroaverage is to calculate the average of the performance metrics for each class independently and includes the following three indicators. Precision is the proportion of the number of correctly classified samples to the number of positive samples judged by the classifier. Recall is

the proportion of correctly classified positive samples to true positive samples. The larger the measurement value is, the stronger the prediction ability of the model would be. The experimental platform of this paper is 64-bit Window 10 system, and the processor is Inter(R) Core (TM) i7-10510U CPU @ 2.30 GHz. The whole experiment is completed with Python 3.7 software.

$$\text{Macro_avg Precision} = \frac{1}{n} \sum_{i=1}^n \text{Precision}_i,$$

Precision_i is the precision of category_i,

$$\text{Macro_avg Recall} = \frac{1}{n} \sum_{i=1}^n \text{Recall}_i, \quad (2)$$

Recall_i is the recall of category_i,

$$\text{Macro_avg F1 - score} = \frac{2 \times \text{Macro_avg Precision} \times \text{Macro_avg Recall}}{\text{Macro_avg Precision} + \text{Macro_avg Recall}}.$$

FIGURE 7: Vibration signals of the sample set under four fault conditions with diameter of 0.007 inches: (a) inner race fault; (b) ball fault; (c) outer race fault; (d) normal.

The parameter setting of each classification model is as follows:

- (1) RF: $n_estimators = 10$, $max_depth = 10$, $criterion = "gini"$, $min_samples_split = 2$, and $min_samples_leaf = 10$
- (2) ET: $n_estimators = 10$, $min_samples_split = 2$, $max_depth = 10$, and $min_samples_leaf = 10$
- (3) XGB: $max_depth = 10$, $learning_rate = 0.1$, and $n_estimators = 10$
- (4) LGBM: $learning_rate = 0.1$, $n_estimators = 10$, and $max_depth = 10$
- (5) SVM: $C:1$, $kernel = "rbf"$, $type = "Classification"$, $gamma = "scale"$, and $tol: 1e-3$
- (6) LSTM: $activation = softmax$, $loss = categorical_crossentropy$, $optimizer = Adam$, $epochs = 500$, and $batch_size = 30$
- (7) DF: multigrained: windows: 300,600,900
 $1*ET + 1*RF$: the same as individual classifiers
 Cascade layer: $2*ET + 2*RF$
 ET, RF: the same as individual classifiers
- (8) Proposed method: multigrained: the same as DF
 Cascade layer: $1*ET + 1*RF + 1*XGB + 1*LGBM$
 ET, RF, XGB, and LGBM: the same as individual classifiers

3.3. Fault Diagnosis and Comparison Analysis. In order to better explain the performance of the improved model, we proposed this study and selected the base model RF, XGBoost, ET, SVM, LightGBM, and long short-term

memory (LSTM) model of deep learning and standard gcForest for comparison. In Experiment 1, rolling bearing faults are divided into four categories for experiments. Table 1 shows the accuracy of these models on the training set and test set. The training accuracy and testing accuracy of the proposed method are 98.72% and 98.54%, respectively. Compared with the other 7 models, the standard deviations of gcForest with the highest and highest accuracy are 1.28% and 1.67%, and the standard deviations of RF with the lowest and highest accuracy are 36.68% and 34.03%, respectively. This fully shows that the recognition accuracy and robustness of the proposed method are greatly improved. The model based on deep forest is much more accurate than all other learning methods, which shows the effectiveness of the ensemble learning model based on tree ensemble. In this experiment, in addition to the SVM, LSTM model accuracy is higher than other accuracy of the shallow learning algorithm due to its strong ability of nonlinear said, but with the SVM, model accuracy is close; it may be that the depth study of the performance of the model parameters of dependence is strong, and parameters setting for each model are described in section 3.2

Figure 8 shows the confusion matrix of the four classification test sets of gcForest and the proposed method. The diagonal elements of the matrix represent the recall rate for each fault mode. It can be seen from Figure 8 that the proposed method in this paper can fully identify the bearing normal condition and outer race fault, and the number of misdiagnoses of inner race fault and ball fault is less than that of gcForest.

F1-score index is the harmonic average of precision and recall, and it is a good comprehensive evaluation index. Figure 9 shows the comparison results of F1 values on the

TABLE 1: Performance of proposed and compared methods under 4 classifications.

Method	Training accuracy (%)	Testing accuracy (%)
RF	62.04	64.51
ET	68.05	69.31
XGBoost	67.02	68.67
LightGBM	78.14	79.33
SVM	81.11	80.38
LSTM	80.81	80.38
gcForest	97.44	96.87
Proposed method	98.72	98.54

FIGURE 8: Confusion matrix comparisons of (a) gcForest and (b) proposed method under 4 classifications.

test set between the proposed method and the basic models in four categories. The F1-score of the method in this paper is consistent with the gcForest under normal conditions but higher than other basic models under other three types of faults, and the F1-score of most basic models is less than 80%. The above results further confirm the performance of the proposed method.

Macroaverage is the average value of each label evaluation calculated independently. The results of the three macroaverage indicators of different methods are shown in Figure 10, which shows that our method achieves the highest value in macroaverage precision, macroaverage recall, and macroaverage F1-score. It is shown that the proposed method has the best effect on four types of fault diagnosis.

In Experiment 2, rolling bearing faults are divided into three different diameters. Together with bearings in normal state, the faults are divided into 10 categories for experiments. The experimental results compared with the basic model are shown in Table 2. Table 2 shows that the accuracy of the deep forest-based learning algorithm is higher than that of other algorithms. Compared with the other 7 models, the proposed method still achieves the highest accuracy. Compared with the gcForest with the highest accuracy in the

base models, the training accuracy and testing accuracy are improved by 2.74% and 2%, respectively.

Figure 11 shows the confusion matrix comparison between the gcForest model and the proposed method under 10 classification test sets. As can be seen from the figure, the proposed method can fully identify six categories, which are NOR, IRF, and ORF with a diameter of 0.007 inches, IRF and ORF with a diameter of 0.021 inches, and IRF with a diameter of 0.014 inches; in addition to the spherical fault misjudgment rate of 0.007 inches in diameter than gcForest slightly higher, the rest of the three conditions of recognition accuracy is higher than that of gcForest, especially the accuracy of ball fault identification increases 0.021 diameter larger extent.

Figure 12 shows the comparison results of F1-score on the test set between the proposed method and the base models in the case of 10 classifications. The F1 value of the method in this paper is high in most types. The best diagnostic results are achieved for all inner race faults, normal condition, and for outer race faults of 0.007 inches and 0.021 inches diameters. For ball faults with diameter of 0.014 and 0.021 and outer race faults with diameter of 0.014 inches, the F1-score is higher than other base models, and only for ball

FIGURE 9: Comparison results of different methods for testing F1-score under 4 classifications.

FIGURE 10: Comparison results of different methods for testing macroaverage under 4 classifications.

faults with diameter of 0.007 inches, it is slightly lower than gcForest, and the F1 value of most base models is around 40%. The above results further demonstrate the superior performance of the proposed method.

Figure 13 shows the three macro average index values of different methods. According to Figure 13, our method achieves the highest value in macroaverage precision, macroaverage recall rate, and macroaverage F1-score. The macro average precision of the proposed method is 57% higher than that of the lowest ET method and 2% higher than that of the highest gcForest. Similarly, the macroaverage recall rate and macroaverage F1-score of the proposed

TABLE 2: Performance of proposed and compared methods under 10 classifications.

Method	Training accuracy (%)	Testing accuracy (%)
RF	37.42	39.10
ET	43.19	40.35
XGBoost	47.68	44.36
LightGBM	55.02	57.89
SVM	67.13	73.43
LSTM	57.65	57.89
gcForest	95.31	94.99
Proposed method	98.05	96.99

method are 56% and 57% higher than those of the lowest method and 1% higher than those of the highest method. It is shown that the proposed method has the best diagnostic performance on 10 types of faults.

In addition, in order to further illustrate the performance of the proposed method, we designed Experiment 3. In Experiment 3, two innovations in the proposed method are analyzed separately for the two classification situations and compared with existing methods in literature [36]. The experimental results are shown in Tables 3 and 4, respectively. Table 3 shows the comparison results of training accuracy and testing accuracy in 4 types of situations, and Table 4 shows the comparison results of training accuracy and testing accuracy in 10 types of situations.

“Only improve the cascade” as mentioned in Section 2.4, a sublevel is added to each cascade layer, which is composed of the full connection of class vectors and placed at the first sublevel of each level. The learner keeps consistent with the original deep forest. “Only replace the learner” means that learners are replaced, and the cascade mode after multi-grained scanning is still consistent with the original deep forest. Tables 3 and 4 show the accuracy of the proposed method is the highest. In Table 4, the accuracy is 98.05% and 96.99%, respectively. It is 4.86% and 1.25% better than the method of only improve the cascade. It is 0.76% and 0.75% better than the method of only replace the learner. The training accuracy and the testing are also higher than

FIGURE 11: Confusion matrix comparisons of (a) gcForest and (b) proposed method under 10 classifications.

FIGURE 12: Comparison results of different methods for testing F1-score under 10 classifications.

FIGURE 13: Comparison results of different methods for testing macroaverage under 10 classifications.

TABLE 3: The results are compared separately with the improved method in 4 classifications.

Method	Training accuracy (%)	Testing accuracy (%)
The literature method	96.72	94.78
Only improve the cascade	96.86	96.03
Only replace the learner	98.06	97.08
Proposed method	98.72	98.54

TABLE 4: The results are compared separately with the improved method in 10 classifications.

Method	Training accuracy (%)	Testing accuracy (%)
The literature method	92.13	91.98
Only improve the cascade	93.19	95.74
Only replace the learner	97.29	96.24
Proposed method	98.05	96.99

existing methods in the literature, which fully proves that the proposed method in our study on the classification result in the rolling bearing fault diagnosis efficiency.

4. Conclusion

This paper proposed an improved method of rolling bearing fault diagnosis based on deep forest. CWRU bearing vibration signals are used to verify the effectiveness of this

method through two different groups of experiments. Connecting multiple scan granularities to add a sublevel at each level of the cascade forest reduces the loss of information flow. In addition, the adoption of more efficient tree model learners not only increases the diversity of classifiers but also helps to improve the recognition rate of fault diagnosis. The analysis results suggest that the detection rates of test faults under the two classifications are 98.54% and 96.99%, respectively, which is higher than those of other base models. The results indicate that the improved deep forest model has high recognition ability and robustness for bearing faults.

This study has a few limitations. Although the proposed method is relatively time-consuming because multiple granularity connections are added for cascading, it can obtain better fault detection performance. In order to solve this problem, feature optimization in the cascade layer is considered in the future research to detect faults efficiently and accurately.

Data Availability

The data of rolling bearing are from the website of Case Western Reserve Lab, and they are all available at <http://csegroups.case.edu/bearingdatacenter/pages/download-data-file>.

Conflicts of Interest

The authors declare that they have no conflicts of interest regarding the publication of this paper.

Acknowledgments

This work was partially supported by the National Natural Science Foundation of China Mathematics Tianyuan Foundation (no. 12026430) and the Scientific Project of Education Department of Jilin Province (no. JJKH20210716KJ). The authors are very grateful for the supports.

References

- [1] G. Xu, D. Hou, H. Qi, and L. Bo, "High-speed train wheel set bearing fault diagnosis and prognostics: a new prognostic model based on extendable useful life," *Mechanical Systems and Signal Processing*, vol. 146, pp. 1–23, 2021.
- [2] T. Xu, Z. Yin, D. Cai, and D. Zheng, "Fault diagnosis for rotating machinery based on local mean decomposition morphology filtering and least square support vector machine," *Journal of Intelligent and Fuzzy Systems*, vol. 32, no. 3, pp. 2061–2070, 2017.
- [3] L. Cui, D. Mo, H. Wang, and P. Chen, "Resonance-based nonlinear demodulation analysis method of rolling bearing fault," *Advances in Mechanical Engineering*, vol. 2013, Article ID 420694, 13 pages, 2013.
- [4] M. Xiao, K. Wen, C. Zhang, X. Zhao, W. Wei, and D. Wu, "Research on fault feature extraction method of rolling bearing based on NMD and wavelet threshold denoising," *Shock and Vibration*, vol. 2018, Article ID 9495265, 11 pages, 2018.

- [5] G. Dong and J. Chen, "Noise resistant time frequency analysis and application in fault diagnosis of rolling element bearings," *Mechanical Systems and Signal Processing*, vol. 33, no. 11, pp. 212–236, 2012.
- [6] W. Xing, Q. Xiangdong, and L. Baojin, "Rolling bearing fault diagnosis based on time-frequency feature parameters and wavelet neural network," *International Journal of Control and Automation*, vol. 8, no. 4, pp. 45–52, 2015.
- [7] H. Helmi and A. Forouzantabar, "Rolling bearing fault detection of electric motor using time domain and frequency domain features extraction and ANFIS," *IET Electric Power Applications*, vol. 13, no. 5, pp. 662–669, 2019.
- [8] G. Yang and Z. Li, "Study on diagnosis algorithm based on wavelet transform for rolling bearing," *Journal of Chemical and Pharmaceutical Research*, vol. 6, no. 7, pp. 2110–2117, 2014.
- [9] N. E. Huang, Z. Shen, S. R. Long et al., "The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis," *Proceedings Mathematical Physical & Engineering Sciences*, vol. 454, no. 1971, pp. 903–995, 1998.
- [10] K. Dragomiretskiy and D. Zosso, "Variational mode decomposition," *IEEE Transactions on Signal Processing*, vol. 62, no. 3, pp. 531–544, 2014.
- [11] L. Y. Zhao, L. Wang, and R. Q. Yan, "Rolling bearing fault diagnosis based on wavelet packet decomposition and multi-scale permutation entropy," *Entropy*, vol. 17, no. 9, pp. 6447–6461, 2015.
- [12] L. Zhang, L. Zhang, J. Hu, and G. Xiong, "Bault diagnosis using a novel classifier ensemble based on lifting wavelet packet transforms and sample entropy," *Shock and Vibration*, vol. 2016, Article ID 4805383, 13 pages, 2016.
- [13] Z. Wu and N. E. Huang, "Ensemble empirical mode decomposition: a noise-assisted data analysis method," *Advances in Adaptive Data Analysis*, vol. 1, no. 1, pp. 1–41, 2009.
- [14] Z. Liu, Y. Liu, H. Shan, B. Cai, and Q. Huang, "Aault diagnosis methodology for gear pump based on EEMD and bayesian network," *PLoS One*, vol. 10, no. 5, 15 pages, Article ID e0125703, 2015.
- [15] X. Qin, Q. Li, X. Dong, and S. Lv, "Tault diagnosis of rolling bearing based on ensemble empirical mode decomposition and random forest," *Shock and Vibration*, vol. 2017, Article ID 2623081, 9 pages, 2017.
- [16] X. Qin, J. Guo, X. Dong, and Y. Guo, "Tault diagnosis of rolling bearing based on variational mode decomposition and iterative random forest," *Shock and Vibration*, vol. 2020, Article ID 1576150, 11 pages, 2020.
- [17] C. Fei, X. Chen, Z. Yang et al., "A rolling bearing fault diagnosis method based on VMD–multiscale fractal dimension/energy and optimized support vector machine," *Journal of Vibroengineering*, vol. 18, no. 6, pp. 3581–3595, 2016.
- [18] D. Hang and J. He, "A novel fault diagnosis method for rolling element bearings using kernel independent component analysis and genetic algorithm optimized RBF network," *Research Journal of Applied Sciences, Engineering and Technology*, vol. 6, no. 5, pp. 895–899, 2013.
- [19] H. Zhao, H. Liu, Y. Jin, X. Dang, and W. Deng, "Feature extraction for data-driven remaining useful life prediction of rolling bearings," *IEEE Transactions on Instrumentation and Measurement*, vol. 70, no. 99, pp. 1–10, 2021.
- [20] W. Deng, S. Shang, X. Cai, H. Zhao, Y. Song, and J. Xu, "An improved differential evolution algorithm and its application in optimization problem," *Soft Computing*, vol. 25, no. 7, pp. 5277–5298, 2021.
- [21] W. Deng, J. Xu, X.-Z. Gao, and H. Zhao, "An enhanced MSIQDE algorithm with novel multiple strategies for global optimization problems," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 99, pp. 1–10, 2020.
- [22] X. Cai, H. Zhao, S. Shang et al., "An improved quantum-inspired cooperative co-evolution algorithm with multi-strategy and its application," *Expert Systems with Applications*, vol. 171, pp. 1–13, Article ID 114629, 2021.
- [23] M. Lv, X. Su, S. Liu et al., "Application of SVM optimized by IPSO in rolling bearing fault diagnosis," *MATEC Web of Conferences*, vol. 227, pp. 1–6, 2018.
- [24] X. Li, W. Zhang, and Q. Ding, "Understanding and improving deep learning-based rolling bearing fault diagnosis with attention mechanism," *Signal Processing*, vol. 161, no. 8, pp. 136–154, 2019.
- [25] J. Feng, Y. Lei, L. Jing et al., "Deep neural networks: a promising tool for fault characteristic mining and intelligent diagnosis of rotating machinery with massive data," *Mechanical Systems and Signal Processing*, vol. 72, no. 73, pp. 303–315, 2016.
- [26] S. Muhammad, K. Cheol-Hong, and K. Jong-Myon, "A hybrid feature model and deep-learning-based bearing fault diagnosis," *Sensors*, vol. 17, no. 12, pp. 1–16, 2017.
- [27] D. T. Hoang and H. J. Kang, "A survey on deep learning based bearing fault diagnosis," *Neurocomputing*, vol. 335, pp. 327–335, 2019.
- [28] F. Wang, H. Jiang, H. Shao et al., "An adaptive deep convolutional neural network for rolling bearing fault diagnosis," *Measurement Science and Technology*, vol. 28, no. 9, pp. 1–17, 2017.
- [29] T. Zhong, J. Qu, X. Fang, H. Li, and Z. Wang, "The intermittent fault diagnosis of analog circuits based on EEMD-DBN," *Neurocomputing*, vol. 436, pp. 74–91, 2021.
- [30] X. Guo, L. Chen, and C. Shen, "Hierarchical adaptive deep convolution neural network and its application to bearing fault diagnosis," *Measurement*, vol. 93, pp. 490–502, 2016.
- [31] G. Xu, M. Liu, Z. Jiang, D. Söffker, and W. Shen, "Bault diagnosis method based on deep convolutional neural network and random forest ensemble learning," *Sensors*, vol. 19, no. 5, pp. 1–21, 2019.
- [32] Z. H. Zhou and J. Feng, "Deep forest: towards an alternative to deep neural networks," in *Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence*, pp. 3553–3559, Melbourne, Australia, August 2017.
- [33] G. Hu, H. Li, Y. Xia, and L. Luo, "A deep Boltzmann machine and multi-grained scanning forest ensemble collaborative method and its application to industrial fault diagnosis," *Computers in Industry*, vol. 100, pp. 287–296, 2018.
- [34] X. Liu, Y. Tian, X. Lei et al., "Deep forest based intelligent fault diagnosis of hydraulic turbine," *Journal of Mechanical Science and Technology*, vol. 33, no. 3, pp. 2049–2058, 2019.
- [35] Z. H. Zhou and J. Feng, "Deep forest," 2017.
- [36] Z. H. Zhou and J. Feng, "Deep forest," *National Science Review*, vol. 6, pp. 1–34, 2019.
- [37] L. Breiman, "Random forests," *Machine Learning*, vol. 45, no. 1, pp. 5–32, 2001.
- [38] P. Geurts, D. Ernst, and L. Wehenkel, "Extremely randomized trees," *Machine Learning*, vol. 63, no. 1, pp. 3–42, 2006.
- [39] Y. Freund and R. E. Schapire, "A decision-theoretic generalization of on-line learning and an application to boosting," *Journal of Computer and System Sciences*, vol. 55, no. 1, pp. 119–139, 1997.

- [40] J. H. Friedman, “Greedy function approximation: a gradient boosting machine,” *The Annals of Statistics*, vol. 29, no. 5, pp. 1189–1232, 2001.
- [41] H. Zeng, C. Yang, H. Zhang et al., “A LightGBM-based EEG analysis method for driver mental states classification,” *Computational Intelligence and Neuroscience*, vol. 2019, Article ID 3761203, 11 pages, 2019.
- [42] H. Wang, Y. Tang, Z. Jia, and F. Ye, “Dense adaptive cascade forest: a self-adaptive deep ensemble for classification problems,” *Soft Computing*, vol. 24, no. 4, pp. 2955–2968, 2020.
- [43] K. A. Loparo, *Bearings Vibration Dataset*, Case Western Reserve University, Cleveland, OH, USA, 2012, <http://csegroups.case.edu/bearingdatacenter/pages/download-data-file>.